AUDIO DECLIPPING BY COSPARSE HARD THRESHOLDING

S. Kitić[†]*, N. Bertin*[⋆]*, R. Gribonval*[†]

[†]Inria , [⋆]IRISA - CNRS UMR 6074

ABSTRACT

Recovery of clipped audio signals is a very challenging inverse problem. Recently, it has been successfully addressed by several methods based on the sparse synthesis data model. In this work we propose an algorithm for enhancement of clipped audio signals that exploits the sparse analysis (*cosparse*) data model. Experiments on real audio data indicate that the algorithm has better signal restoration performance than state-of-the-art sparse synthesis declipping methods.

***Index Terms*—** clipping, audio, sparse, cosparse, inverse problems

1. INTRODUCTION

Audio signals are prone to various degradations and clipping is among the most common ones. It may arise during recording due to the poor dynamic range of a microphone, or of analog-to-digital circuits, but it may be seen also as an extreme case of magnitude compression [1].

Clipping inevitably causes an information loss: in the time domain, the signal is "cut-off" when its amplitude is above some threshold. This manifests as an expansion (spread) in the frequency domain. An intuitive approach to de-clip clipped signals may be to "clean" their frequency spectrum and preserve only those frequencies needed to represent the original signal accurately. However, estimation of the correct support in the frequency domain is also a difficult problem. Hence, we need some additional prior information about the underlying signal.

We formalize clipping as the following element-wise operation on the original signal $\boldsymbol{x} \in \mathbb{R}^n$ (yielding the clipped signal $\bar{\boldsymbol{x}}$):

$$\bar{\boldsymbol{x}} = \operatorname{sign}(\boldsymbol{x}) \min(|\boldsymbol{x}|, \tau) \tag{1}$$

The threshold τ represents amplitude saturation level (here we assume *symmetric* clipping, *i.e.* the same threshold is applied to both positive and negative samples of $\boldsymbol{x}$).

Let $\mathcal{M}_{\mathrm{r}}(\bar{\boldsymbol{x}})$ be the sampling operator which extracts unclipped, "*reliable*", samples, and $\mathcal{M}_{\mathrm{c}}^{+}(\bar{\boldsymbol{x}})$, $\mathcal{M}_{\mathrm{c}}^{-}(\bar{\boldsymbol{x}})$ be the operators which extract clipped samples with positive and negative amplitude, respectively. Knowing that (for the clipped positive/negative samples) the amplitude of the original signal has to be above (respectively, below) the threshold, we can formulate the inverse problem of signal de-clipping in the following way:

$$\begin{aligned} \textit{Find } \hat{\boldsymbol{x}} \in \mathbb{R}^n \textit{ such that: } \mathcal{M}_{\mathrm{r}}(\hat{\boldsymbol{x}}) &= \mathcal{M}_{\mathrm{r}}(\bar{\boldsymbol{x}}) \\ \mathcal{M}_{\mathrm{c}}^{+}(\hat{\boldsymbol{x}}) &\geq \mathcal{M}_{\mathrm{c}}^{+}(\bar{\boldsymbol{x}}) \\ \mathcal{M}_{\mathrm{c}}^{-}(\hat{\boldsymbol{x}}) &\leq \mathcal{M}_{\mathrm{c}}^{-}(\bar{\boldsymbol{x}}) \end{aligned} \tag{2}$$

This is an ill-posed inverse problem, since there are infinitely many possible solutions for $\hat{\boldsymbol{x}}$. We chose to regularize the problem by introducing a *sparsity hypothesis*.

2. SPARSE ANALYSIS DATA MODEL

The premise of sparse regularization is that many signals have low-dimensional representation, but not necessarily in their original domain. This assumption is sometimes sufficient to mitigate ill-posedness of the initial problem.

The *sparse analysis*, also known as the *cosparse* data model assumes that the signal in question can be "sparsified" by applying an adequate linear transform. Let $\boldsymbol{\Omega} \in \mathbb{R}^{p \times n}, (p \geq n)$ denote the matrix form of this sparsifying transform and let the linear constraints in (2) be denoted by Γ. We introduce the following optimization problem:

$$\operatorname{minimize}_{\hat{\boldsymbol{x}}} \|\boldsymbol{\Omega}\hat{\boldsymbol{x}}\|_0 \text{ s.t. } \hat{\boldsymbol{x}} \text{ respects } \Gamma \tag{3}$$

However, minimizing this ℓ_0-"norm" is NP-hard, but there are convex relaxation and greedy algorithms [2] which may be used to approximate its solution.

The cosparse data model has only recently attracted the attention of scientific community, and to our best knowledge, it has not yet been exploited to solve the declipping problem.

3. COSPARSE DECLIPPING

Our approach is based on transposing the *Consistent Iterative Hard Thresholding* (Consistent IHT) algorithm [3] to the cosparse data model. Consistent IHT relies on applying hard

This work was supported in part by the European Research Council, PLEASE project (ERC-StG-2011-277906).

[†]Centre Inria Rennes - Bretagne Atlantique, France

[⋆]Rennes, France

thresholding operator $H_K(\boldsymbol{s})$ [1] to the transform coefficient vector $\boldsymbol{s}$ in order to approximate sparse solutions. The transform coefficients are estimated by performing a gradient descent step of an objective function that incorporates the constraints expressed in Γ.

A pragmatical issue is that computing the k-cosparse projection is proven to be NP-hard [4]. To circumvent this problem, we take a different approach, largely based on the *Alternating Direction Method of Multipliers* [5] framework. The resulting algorithm is presented as pseudocode in algorithm 1 and will be referred to as Cosparse Declipping by Hard Thresholding (CoDec-HT).

Algorithm 1 COsparse DEClipping by Hard Thresholding

Require: $\boldsymbol{\Omega}, \bar{\boldsymbol{x}}, \Gamma, \varepsilon$

1: $\hat{\boldsymbol{x}}^{(0)} = \bar{\boldsymbol{x}}, \boldsymbol{u}^{(0)} = \boldsymbol{0}, k = 1$
2: $\hat{\boldsymbol{z}}^{(k)} = H_k\left(\boldsymbol{\Omega}\hat{\boldsymbol{x}}^{(k-1)} + \boldsymbol{u}^{(k-1)}\right)$
3: $\hat{\boldsymbol{x}}^{(k)} = \arg\min_{\tilde{\boldsymbol{x}}} \|\boldsymbol{\Omega}\tilde{\boldsymbol{x}} - \hat{\boldsymbol{z}}^{(k)} + \boldsymbol{u}^{(k-1)}\|_2^2$ s.t. $\tilde{\boldsymbol{x}}$ respects Γ
4: **if** $\|\boldsymbol{\Omega}\hat{\boldsymbol{x}}^{(k)} - \hat{\boldsymbol{z}}^{(k)} + \boldsymbol{u}^{(k-1)}\|_\infty \leq \varepsilon$ **then**
5: terminate
6: **else**
7: $\boldsymbol{u}^{(k)} = \boldsymbol{u}^{(k-1)} + \boldsymbol{\Omega}\hat{\boldsymbol{x}}^{(k)} - \hat{\boldsymbol{z}}^{(k)}$
8: $k \leftarrow k + 1$
9: go to 2
10: **end if**
11: **return** $\hat{\boldsymbol{x}}^{(k)}$

Mimicking the heuristics used in Consistent IHT, we relax the sparsity k of $\boldsymbol{z}$ through iterations, which allows to "learn" the unknown sparsity level of a given signal. Through iterations, the vectors $\boldsymbol{\Omega}\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{z}}$ eventually get close to each other, thus the signal estimate $\hat{\boldsymbol{x}}$ becomes approximately cosparse.

The analysis operator $\boldsymbol{\Omega}$ can be chosen as any (possibly overcomplete) transform which sparsifies the audio signal. In our experiments we used concatenated Discrete Cosine Transform and Discrete Sine Transform matrices.

4. EXPERIMENTS

We carried declipping experiments with 10-second excerpts from two wideband music pieces, sampled at 16kHz with 16bit encoding. Audio 1 is a rock piece with vocals, drums, two electric guitars and bass, while Audio 2 is an instrumental piece performed on an analogue synthesizer. The clipped versions of these signals are divided into overlapping frames of length 1024 before processing. Afterwards, the outputs are re-synthesized by the overlap and add scheme.

Beforehand, the audio is clipped given a predefined SNR_{inp} (signal-to-noise ratio of the input) value. The restoration performance is measured as the output signal-to-noise ratio (SNR_{out}):

$$\text{SNR}_{\text{inp}} = 20\log_{10}\frac{\|\boldsymbol{x}\|_2}{\|\boldsymbol{x} - \bar{\boldsymbol{x}}\|_2}, \text{SNR}_{\text{out}} = 20\log_{10}\frac{\|\boldsymbol{x}\|_2}{\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2} \tag{4}$$

The proposed algorithm is compared against [3] and [6][2], which serve as a reference. Consistent IHT (with dictionary $\boldsymbol{\Psi} = \boldsymbol{\Omega}^H$) and the proposed algorithm are based on a similar algorithmic methodology and differ only by the underlying data model. Algorithm [6] exploits a structured prior termed *social sparsity* [7].

The results[3] presented in Figure 1 suggest that cosparsity-based hard thresholding outperforms synthesis-based one at all clipping levels. Interestingly, the SNR improvement rate of CoDec HT is roughly of the same order as for the social sparsity based algorithm, despite the fact that the latter uses a stronger prior.

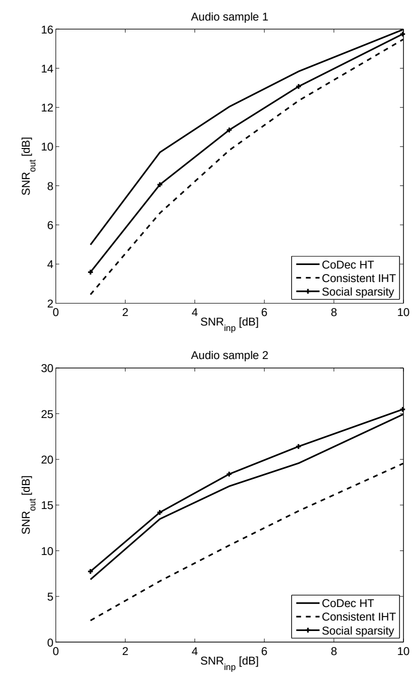

Fig. 1: Declipping results for two wideband audio tracks

5. CONCLUSION

Experimental results indicate that sparse analysis regularization can be successfully applied to the audio declipping inverse problem. Future work will aim at incorporating structured cosparsity in the data model to improve declipping performance, and at addressing extended magnitude corruption scenarios, such as dynamic range compression or soft clipping.

[1] $H_K(\boldsymbol{s})$ keeps K highest in absolute magnitude elements of $\boldsymbol{s}$ and sets the remaining to zero.

[2] We acknowledge M.Kowalski who provided his original code.

[3] Publicly available at <http://people.rennes.inria.fr/Srdan.Kitic/>.

6. REFERENCES

[1] J. O. Smith, *Physical audio signal processing: For virtual musical instruments and audio effects*, W3K Publishing, 2010.

[2] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "The Cosparse Analysis Model and Algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.

[3] S. Kitić, L. Jacques, N. Madhu, M. P. Hopwood, A. Spriet, and C. De Vleeschouwer, "Consistent iterative hard thresholding for signal declipping," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013.

[4] A. M. Tillmann, R. Gribonval, and M. E. Pfetsch, "Projection onto the k-cosparse set is NP-hard," *CoRR*, vol. abs/1303.5305, 2013.

[5] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[6] K. Siedenburg, M. Kowalski, and M. Dörfler, "Audio declipping with social sparsity," *Submitted to ICASSP, 2014*, 2013.

[7] M. Kowalski, K. Siedenburg, and M. Dörfler, "Social sparsity! neighborhood systems enrich structured shrinkage operators," 2012.